

Infinitely exchangeable random graphs generated from a Poisson point process on monotone sets and applications to cluster analysis for networks

Harry Crane

Abstract

We construct an infinitely exchangeable process on the set \mathcal{E} of subsets of the power set of the natural numbers \mathbb{N} via a Poisson point process with mean measure Λ on the power set of \mathbb{N} . Each $E \in \mathcal{E}$ has a least monotone cover in \mathcal{F} , the collection of monotone subsets of \mathcal{E} , and every monotone subset maps to an undirected graph $G \in \mathcal{G}$, the space of undirected graphs with vertex set \mathbb{N} . We show a natural mapping $\mathcal{E} \rightarrow \mathcal{F} \rightarrow \mathcal{G}$ which induces an infinitely exchangeable measure on the projective system $\mathcal{G}^{\text{rest}}$ of graphs \mathcal{G} under permutation and restriction mappings given an infinitely exchangeable family of measures on the projective system $\mathcal{E}^{\text{rest}}$ of subsets with permutation and restriction maps. We show potential connections of this process to applications in cluster analysis, machine learning, classification and Bayesian inference.

1 Introduction

Here, we show a construction of an infinitely exchangeable family of random graphs which is based on an associated Poisson point process on the power set of the natural numbers \mathbb{N} . We provide a necessary and sufficient condition for the induced random graph to be infinitely exchangeable and discuss a potential use for this model in the area of cluster analysis and stochastic classification, which has been previously studied in a statistical and machine learning context in previous work by Jordan [2] (with Blei and Ng) and [7] (with Broderick and Pitman) McCullagh [10, 11], but outside of the realm of network analysis.

We now introduce preliminary material and notation which is critical to our treatment.

1.1 Projective systems

A *projective system* associates with each finite set $[n]$ a set Q_n and with each one-to-one injective map $\varphi : [m] \rightarrow [n]$, $m \leq n$, a projection $\varphi^* : Q_n \rightarrow Q_m$ which maps Q_n into Q_m such that

- if φ is the identity $[n] \rightarrow [n]$ then φ^* is the identity $Q_n \rightarrow Q_n$ and
- if $\psi : [l] \rightarrow [m]$, $l \leq m$, and $\psi^* : Q_m \rightarrow Q_l$ is its associated projection, then the composition $(\varphi\psi) : [l] \rightarrow [n]$ satisfies $(\varphi\psi)^* \equiv \psi^*\varphi^* : Q_n \rightarrow Q_l$.

If Q_n is the set of subsets of $[n]^2$, i.e. the space of directed graphs with n vertices, one can define the projection $Q_n \rightarrow Q_m$ either by *restriction* or *delete-and-repair*. Each $A \in Q_n$ can be represented as an

$n \times n$ matrix with entries in $\{0, 1\}$ such that $A_{ij} = 1$ if $(i, j) \in A$ and $A_{ij} = 0$ otherwise. For each $n \geq 1$, let $\varphi_{n,n+1}$ be the operation on Q_{n+1} which restricts A to the complement of $\{n+1\}$. In matrix form, $\varphi_{n,n+1}A =: A|_{[n]}$ is the $n \times n$ matrix obtained from A by removing the last row and last column of A and keeping the rest of the entries unchanged. It is clear that the compositions $\varphi_{m,n} := \varphi_{m,m+1} \circ \cdots \circ \varphi_{n-1,n}$ for $m \leq n$ are well-defined as the restriction of $A \in Q_n$ to $[m]$ by removing the last $n-m$ rows and columns of A .

For $n \geq 1$, we write \mathcal{S}_n to denote the symmetric group of permutations of $[n]$, i.e. one-to-one maps $[n] \rightarrow [n]$. Each $\sigma \in \mathcal{S}_n$ acts on each element $A \in Q_n$ componentwise in the usual way. That is, $(i, j) \in A$ if and only if $(\sigma(i), \sigma(j)) \in \sigma(A)$. The restriction maps $(\varphi_{m,n}, m \leq n)$ together with permutation maps $(\sigma \in \mathcal{S}_n, n \geq 1)$ and their compositions make $Q := (Q_n, n \geq 1)$ a projective system.

Another way to specify a projective system on $(Q_n, n \geq 1)$ is by *delete-and-repair*. For $n \geq m \geq 1$, let ψ_m act on $A \in Q_n$ by removing the m th row and column of A and directing an edge from each i in $\{j \in [n] : (j, m) \in A\}$ to each k in $\{j \in [n] : (m, j) \in A\}$. In other words, $\psi_m A$ is obtained by deleting the vertex labeled m from A and connecting two vertices i and k by a directed edge from i to k if both (i, m) and (m, k) are elements of A , i.e. there is a directed path $i \rightarrow m \rightarrow k$ in A .

For $m \leq n$, define $\psi_{m,n} := \psi_{m+1} \circ \cdots \circ \psi_n$. Plainly, $\psi_{m,n}$ is well-defined since for each $n \geq 2$, $\psi_{n-2,n} \equiv \psi_{n-1} \circ \psi_n = \psi_n \circ \psi_{n-1}$ and $\psi_{l,n} = \psi_{l,m} \circ \psi_{m,n}$. The delete-and-repair maps $(\psi_{m,n}, m \leq n)$ together with permutation maps $(\sigma \in \mathcal{S}_n, n \geq 1)$ and compositions also make $(Q_n, n \geq 1)$ a projective system, which differs from the above projective system based on restriction maps.

Throughout the rest of this paper, for a system $Q := (Q_n, n \geq 1)$ we write $\varphi_{m,n}^Q$ to represent the restriction maps for collection Q and $\psi_{m,n}^Q$ to represent the corresponding delete-and-repair maps. We write Q^{rest} , resp. Q^{dr} , to denote the projective system on Q described by the restriction, resp. delete-and-repair, maps together with permutation maps.

1.2 Power set of $2^{[n]}$

Let $\mathcal{E}_n = 2^{2^{[n]}}$ denote the power set of all subsets of $2^{[n]}$, the power set of $[n]$. For each $n \geq 1$, we define the partial order $\leq_{\mathcal{E}}$ on \mathcal{E}_n by

$$E \leq_{\mathcal{E}} E' \iff [e \in E \Rightarrow e \in E'].$$

Therefore, $\{\{1\}\} \leq_{\mathcal{E}} \{\{1\}, \{1, 2\}\}$ but $\{\{1\}\}$ and $\{\{1, 2\}\}$ have no relation under $\leq_{\mathcal{E}}$.

We define restriction maps $\varphi_{m,n}^{\mathcal{E}} : \mathcal{E}_n \rightarrow \mathcal{E}_m$ as follows. For $E := \{e_i : 1 \leq i \leq k\} \in \mathcal{E}_n$, we define $\varphi_{m,n}^{\mathcal{E}}(E) := E|_{[m]} := \{e_i \cap [m] : 1 \leq i \leq k\}$. The maps $(\varphi_{m,n}^{\mathcal{E}}, m \leq n)$ preserve $\leq_{\mathcal{E}}$ on \mathcal{E} since

$$E \leq_{\mathcal{E}} E' \iff [e \in E \implies e \in E']$$

and so

$$e \in E \implies e \cap [m] \in \varphi_{m,n}^{\mathcal{E}} E, e \cap [m] \in \varphi_{m,n}^{\mathcal{E}} E' \implies \varphi_{m,n}^{\mathcal{E}} E \leq_{\mathcal{E}} \varphi_{m,n}^{\mathcal{E}} E'.$$

Permutations act on elements of E componentwise, i.e. for $n \geq 1, E \in \mathcal{E}_n$, and $\sigma \in \mathcal{S}_n$, we have $\sigma(E) := \{\sigma(e) : e \in E\}$. We write $\mathcal{E}^{\text{rest}}$ to denote the projective system $\mathcal{E} := (\mathcal{E}_n, n \geq 1)$ together with restriction maps $(\varphi_{m,n}^{\mathcal{E}}, m \leq n)$ and permutation maps $(\sigma \in \mathcal{S}_n, n \geq 1)$ and their compositions.

1.3 Monotone sets

A subset $A \subset 2^{[n]}$ is *monotone* if $a \in A$ implies $2^a \subset A$. For example, the set $A = \{\emptyset, \{1\}, \{2\}, \{3\}, \{1, 2\}\} = \langle \{1, 2\}, \{3\} \rangle$ is a monotone set with maximal elements $\{1, 2\}$ and $\{3\}$, which constitute the generating class of A , written $G(A)$. An element a of the generating class of a monotone set A is a maximal element of A in the sense that no other element $a' \in A$ contains a as a subset. The generating class $G(A)$ of a monotone subset A consists of all maximal elements of A . We write \mathcal{F}_n as the set of monotone sets taking elements in $2^{[n]}$. Note that a monotone set is uniquely determined by its generating class, and so we will write $A \in \mathcal{F}_n$ and $G(A)$ to describe the same object, i.e. the monotone set A , whenever it is convenient to do so. Every subset E of $2^{[n]}$ has a least monotone cover in \mathcal{F}_n , which we denote by $\alpha(E)$ and is given by $\alpha(E) := \{2^a : a \in E\}$.

For each $n \geq 1$, \mathcal{F}_n is a partially ordered set induced by the partial order *inclusion* on $2^{[n]}$, i.e. for $A, B \in \mathcal{F}_n$, we say $A \leq_{\mathcal{F}} B$ if and only if each $a \in G(A)$ is a subset of some $b \in G(B)$. And for any pair A, B with $A \leq B$, the intervals $[A, B]$, $(A, B]$ and $[A, B)$ are well-defined subsets of \mathcal{F}_n . Note that we intend the symbols \subseteq and \subset to have strictly different meanings in this paper. In particular, we write $A \subseteq B$ to mean A is any subset of B , while we write $A \subset B$ to mean A is a *proper subset* of B , i.e. $A \subseteq B$ but $A \neq B$. This distinction becomes important in the next section.

We define the operation *restriction* $\varphi_{n,n+1}^{\mathcal{F}} : \mathcal{F}_{n+1} \rightarrow \mathcal{F}_n$ as the operation which maps $G(A) \mapsto G(A) \cap [n]$. That is, for $A \in \mathcal{F}_{n+1}$ with $G(A) := \{A_1, \dots, A_k\}$, $\varphi_{n,n+1}^{\mathcal{F}} A := \langle A_i \cap [n] : i = 1, \dots, k \rangle$. For $m \leq n$, we define $\varphi_{m,n}^{\mathcal{F}} := \varphi_{m,m+1}^{\mathcal{F}} \circ \dots \circ \varphi_{n-1,n}^{\mathcal{F}}$ in the usual way by composition, and the collection of restriction maps $\mathcal{F} := (\varphi_{m,n}^{\mathcal{F}}, m \leq n)$ together with permutation maps makes $(\mathcal{F}_n, n \geq 1)$ into a projective system, written $\mathcal{F}^{\text{rest}}$. Here, a permutation $\sigma \in \mathcal{S}_n$ acts on a monotone set $A \in \mathcal{F}_n$ by acting componentwise on its generating class, i.e. $a \in G(A)$ if and only if $\sigma(a) \in G(\sigma(A))$.

Also note that the inverse mapping $\varphi_{n,n+1}^{\mathcal{F}-1}$ associates with each $F \in \mathcal{F}_n$ an interval, and also maps intervals in \mathcal{F}_n to intervals in \mathcal{F}_{n+1} . In particular, for $n \geq 1$ and $A := \langle A_1, \dots, A_k \rangle \in \mathcal{F}_n$, we have

$$\varphi_{n,n+1}^{\mathcal{F}-1}(A) = [\langle G(A) \rangle, \langle A_j \cup \{n+1\} : j = 1, \dots, k \rangle].$$

1.4 Undirected graphs

For $n \geq 1$, an *undirected graph* $G \in \mathcal{G}_n$ is a pair (V, E) of vertices V and edges E whereby the number of vertices $\#V$ of G is n , and without loss of generality we assume $V = [n] := \{1, \dots, n\}$, and the edges are a subset of $[n]^2$ such that $(i, j) \in E$ implies $(j, i) \in E$.

For each $n \geq 1$, the elements of \mathcal{G}_n correspond to the symmetric subsets of $[n]^2$, e.g. for $A \in \mathcal{G}_n$, $(i, j) \in A$ if and only if $(j, i) \in A$, and hence $\mathcal{G} := (\mathcal{G}_n, n \geq 1)$ is a projective system under both restriction and delete-and-repair, as described in section 1.1. For each $n \geq 1$, \mathcal{G}_n is a poset under $\leq_{\mathcal{G}}$ where for $G, G' \in \mathcal{G}_n$

$$G \leq_{\mathcal{G}} G' \iff [\{i, j\} \in G \implies \{i, j\} \in G'].$$

For clarity of notation, write $\varphi_{m,n}^{\mathcal{G}}$ and $\psi_{m,n}^{\mathcal{G}}$ to denote the operations of restriction and delete-and-repair respectively on the projective system $\mathcal{G}^{\text{rest}}$ and \mathcal{G}^{dr} respectively.

1.5 Some category theory

The relationship between the collections \mathcal{E} , \mathcal{F} , and \mathcal{G} is described in a straightforward way by elementary concepts in category theory [1].

A *category* \mathbf{C} consists of *objects* A, B, C, \dots and *arrows* f, g, h, \dots between objects so that for each arrow $f \in \mathbf{C}$ there are objects $\text{dom}(f)$ and $\text{cod}(f)$ in \mathbf{C} , the domain and codomain respectively of f , and we write $f : \text{dom}(f) \rightarrow \text{cod}(f)$. Given $f, g \in \mathbf{C}$ such that $\text{cod}(f) = \text{dom}(g)$, the *composite* $g \circ f : \text{dom}(f) \rightarrow \text{cod}(g)$ is an arrow in \mathbf{C} . Also, for each object $A \in \mathbf{C}$ there is an identity arrow $1_A : A \rightarrow A$, and all arrows of \mathbf{C} must satisfy associativity and preservation under composition with the identity functions.

In each of the categories we define below, there is at most one arrow between any two objects. Therefore, if an arrow f corresponds to the composition of arrows $g \circ h$, these must represent the same arrow. Under this assumption, for a pair of objects we need not make explicit all of the various compositions of arrows which result in an arrow between these objects, as they are implicitly assumed to be there, and are all the same arrow.

We define three categories as follows. We write $\mathcal{E}^{\text{rest}}$ for the category with objects given by the elements of \mathcal{E} and arrows given by the restriction $(\varphi_{m,n}^{\mathcal{E}}, m \leq n)$ and permutation $(\sigma \in \mathcal{S}_n, n \geq 1)$ maps, and compositions of these maps. That is, for $m \leq n$ and $E \in \mathcal{E}_n$, there is an arrow $E \rightarrow \varphi_{m,n}^{\mathcal{E}} E = E|_{[m]}$. We denote the arrow between E and $E|_{[m]}$ by $\varphi_{m,n}^{\mathcal{E}} E$. Likewise, we define $\mathcal{F}^{\text{rest}}$ and $\mathcal{G}^{\text{rest}}$ to be the category with objects given by the elements of \mathcal{F} and \mathcal{G} respectively, and arrows defined by the restriction maps, $(\varphi_{m,n}^{\mathcal{F}}, m \leq n)$ and $(\varphi_{m,n}^{\mathcal{G}}, m \leq n)$ resp., and permutation maps. For example, in $\mathcal{F}^{\text{rest}}$, we have $\langle \{1, 2\}, \{3\} \rangle \rightarrow_{\varphi_{2,3}^{\mathcal{F}}} \langle \{1, 2\} \rangle$.

A *functor* between categories \mathbf{C} and \mathbf{D} is a map $F : \mathbf{C} \rightarrow \mathbf{D}$ which takes objects in \mathbf{C} to objects in \mathbf{D} and arrows in \mathbf{C} to arrows in \mathbf{D} . There are natural functors $\alpha : \mathcal{E}^{\text{rest}} \rightarrow \mathcal{F}^{\text{rest}}$ and $\beta : \mathcal{F}^{\text{rest}} \rightarrow \mathcal{G}^{\text{rest}}$, which we define as follows.

- (i) For each $n \geq 1$ and $E \in \mathcal{E}_n$, $\alpha(E) \in \mathcal{F}_n$ is the least monotone cover of E .
- (ii) For each $n \geq 1$ and $F := \langle F_1, \dots, F_k \rangle \in \mathcal{F}_n$, $\beta(F) = G_F \in \mathcal{G}_n$ where $(i, j) \in G_F$ if and only if $\{i, j\} \subseteq F_j$ for some $j = 1, \dots, k$.
- (iii) For $m \leq n$, $\alpha(\varphi_{m,n}^{\mathcal{E}}) = \varphi_{m,n}^{\mathcal{F}}$ and $\beta(\varphi_{m,n}^{\mathcal{F}}) = \varphi_{m,n}^{\mathcal{G}}$ so that $(\alpha \circ \beta)(\varphi_{m,n}^{\mathcal{E}}) = \varphi_{m,n}^{\mathcal{G}}$.
- (iv) For every $n \geq 1$, $\alpha(\sigma) = \sigma$ and $\beta(\sigma) = \sigma$ for all $\sigma \in \mathcal{S}_n$.

Functors between posets preserve partial ordering. For $\bullet = \mathcal{E}, \mathcal{F}, \mathcal{G}$, we have defined the partial order \leq_{\bullet} in previous sections. For $m \leq n$ and $A, A' \in \bullet_n$ such that $A \leq_{\bullet} A'$, it holds that $\varphi_{m,n}^{\bullet} A \leq_{\bullet} \varphi_{m,n}^{\bullet} A'$. Hence, for each fixed $1 \leq m \leq n$, the restriction maps $(\varphi_{m,n}^{\bullet})$ define a functor $\bullet_n \rightarrow \bullet_m$, where we take the elements of \bullet_n and \bullet_m as objects and the partial order \leq_{\bullet} defines the arrows, i.e. there is an arrow $A \rightarrow A'$ in \bullet_n if and only if $A \leq_{\bullet} A'$.

We can also regard the collections $\mathcal{E}^{\text{rest}}, \mathcal{F}^{\text{rest}}, \mathcal{G}^{\text{rest}}$ as partially ordered sets with partial order $\leq_{\bullet}^{\text{rest}}$, $\bullet = \mathcal{E}, \mathcal{F}, \mathcal{G}$, defined as follows. For $m \leq n$, $A \in \bullet_n$, $A' \in \bullet_m$,

$$A' \leq_{\bullet}^{\text{rest}} A \iff A' = \varphi_{m,n}^{\bullet} A.$$

The functors α, β and $\beta \circ \alpha$ above preserve the partial orders $\leq_{\mathcal{E}^{\text{rest}}}, \leq_{\mathcal{F}^{\text{rest}}}$ and $\leq_{\mathcal{G}^{\text{rest}}}$.

Below, we show a construction of infinitely exchangeable random elements of $\mathcal{E}^{\text{rest}}$. In particular, we construct an infinitely exchangeable random monotone set by projecting from a Poisson point process X on the power set to a random subset $X^* \in \mathcal{E}^{\text{rest}}$ of the power set to the least monotone cover $\alpha(X^*) \in \mathcal{F}^{\text{rest}}$, which corresponds to a random graph $\beta \circ \alpha(X^*) \in \mathcal{G}^{\text{rest}}$. This procedure looks like this

$$X \rightarrow_* X^* \rightarrow_\alpha \alpha(X^*) \rightarrow_\beta \beta \circ \alpha(X^*).$$

Projective systems in statistics The relevance of category theory and projective systems in statistical modeling is discussed in detail by McCullagh [9]. Here we have introduced projective systems for the collection of subsets of $2^{[n]}$, their associated monotone subsets of $2^{[n]}$ and their associated undirected graphs in $[n]^2$.

The choice of projection, i.e. either restriction or delete-and-repair, on \mathcal{G} is intended to reflect the notion of subsampling in statistics, and each admits its own statistical interpretation in terms of subsampling which may be appropriate depending on the application, and the actual way in which observations are made. In the study of directed graphs, in particular permutations, delete-and-repair is often used, see e.g. the Chinese restaurant process (CRP) on permutations [12]. However, for our purposes, restriction maps are a natural choice.

To be specific, if we think of a graph as modeling a social network, the edges of the graph represent social links among the agents (nodes) in the population. When observing a social network, there are several sampling methods which are reasonable. The most intuitive of these are perhaps node sampling and snowball sampling. In node sampling, we first sample a set $\{u_1, \dots, u_n\}$ of nodes and subsequently observe any edges between these nodes. In this setting, restriction accurately reflects our sampling method since removal of a node from our sample, e.g. u_n , removes any edges which involves that node from our sampled graph. In snowball sampling, we start with a set of nodes, without loss of generality suppose we start with one node u , and proceed as follows. Given $U_0 = u$, sample $U_1 := \{v \in \mathcal{U} : v \sim U_0\}$, the set of nodes v adjacent to some node in U_0 in the graph we are observing. Subsequently, put $U_{k+1} := \{v \in \mathcal{U} : v \sim U_k, v \notin U_k, U_{k-1}, \dots, U_0\}$. In snowball sampling, we stop at some arbitrary level k . This way of sampling results in a sample of all nodes at radius k or less from u , and depends on our choice of u and the network structure. In this case, removal of a node $v \in V := \bigcup_{i=0}^k U_i$ results in removal of all other nodes $w \in V$ such that v lies along every path between u and w . The result of this type of subsampling is not described simply via restriction. Below, we discuss only those projective systems characterized by the restriction maps, and the implications of this method of sampling on inference.

1.6 Infinite exchangeability

A family $(p_n, n \geq 1)$ of probability measures on a projective system $Q := (Q_n, n \geq 1)$ is *infinitely exchangeable* if it is invariant under both the action of permutations, called *finite exchangeability*, and selection according to the projection maps $(D_{m,n}, m \leq n)$ associated with the system, called *consistency*. For example, a family of measures $(p_n, n \geq 1)$ on Q is infinitely exchangeable if

- for each $n \geq 1$ and $\sigma \in \mathcal{S}_n$, $p_n(A) = p_n(\sigma(A))$ for every $A \in Q_n$ and
- for every $m \leq n$, $p_m(A) = p_n(D_{m,n}^{-1}(A))$ for every $A \in Q_m$.

An infinitely exchangeable collection of measures uniquely characterizes a measure p on the infinite space associated with Q through its finite-dimensional distributions and invariance under projection and permutation maps.

Above we have defined functors $\alpha : \mathcal{E}^{\text{rest}} \rightarrow \mathcal{F}^{\text{rest}}$ and $\beta : \mathcal{F}^{\text{rest}} \rightarrow \mathcal{G}^{\text{rest}}$. One property of functors is that $\alpha(E_1 \circ E_2) = \alpha(E_1) \circ \alpha(E_2)$. Hence, we have the following elementary lemma which is useful for our purposes below.

Lemma 1.1. *Let $\gamma : \mathbf{C} \rightarrow \mathbf{D}$ be a functor between categories \mathbf{C} and \mathbf{D} where the objects C of \mathbf{C} and D of \mathbf{D} form a projective system under mappings $\pi^{\mathbf{C}}$ and $\pi^{\mathbf{D}}$ respectively, and the arrows are defined by the partial ordering induced by the projections $\pi^{\mathbf{C}}$ and $\pi^{\mathbf{D}}$. Let $\mu_{\mathbf{C}}$ be a probability measure on C and let $\mu_{\mathbf{D}} = \mu_{\mathbf{C}}\gamma^{-1}$ be the distribution induced on the objects of \mathbf{D} by $\mu_{\mathbf{C}}$ through the functor γ . If $\mu_{\mathbf{C}}$ is invariant under $\pi^{\mathbf{C}}$, i.e. $\mu_{\mathbf{C}} = \mu_{\mathbf{C}}\pi^{\mathbf{C}^{-1}}$, then $\mu_{\mathbf{D}}$ is invariant under $\pi^{\mathbf{D}}$.*

Proof. Since $\gamma : \mathbf{C} \rightarrow \mathbf{D}$ is a functor, we have $\gamma(g \circ f) = \gamma(g) \circ \gamma(f)$ for any arrows $f, g \in \mathbf{C}$ such that $\text{dom}(g) = \text{cod}(f)$. Let c be an object in \mathbf{C} and $c' := \pi^{\mathbf{C}}(c)$, so that we have an arrow $c \rightarrow_{\pi^{\mathbf{C}}(c)} c'$ which we denote by $\pi^{\mathbf{C}}(c)$, and the action of the functor γ is such that

$$(\gamma \circ \pi^{\mathbf{C}})(c) = \gamma(\pi^{\mathbf{C}}(c)) = \pi^{\mathbf{D}}(\gamma(c')) = (\pi^{\mathbf{D}} \circ \gamma)(c'),$$

and we have $\gamma \circ \pi^{\mathbf{C}} = \pi^{\mathbf{D}} \circ \gamma$ and invariance of the induced measure $\mu_{\mathbf{D}}$ follows. Indeed, for $d = \gamma(c)$ and $d' = \gamma(c')$,

$$\mu_{\mathbf{D}}(d') \equiv \mu_{\mathbf{C}}\gamma^{-1}(d') = (\mu_{\mathbf{C}}\pi^{\mathbf{C}^{-1}})\gamma^{-1}(d') = \mu_{\mathbf{C}}(\gamma \circ \pi^{\mathbf{C}})^{-1}(d') = \mu_{\mathbf{C}}(\pi^{\mathbf{D}} \circ \gamma)^{-1}(d') = (\mu_{\mathbf{C}}\gamma^{-1})\pi^{\mathbf{D}^{-1}}(d') = \mu_{\mathbf{D}}\pi^{\mathbf{D}^{-1}}(d').$$

□

This will simplify our proofs below for infinite exchangeability of the random graph induced by a random subset of $2^{[n]}$.

2 Construction of an infinitely exchangeable random graph

Let X be a Poisson point process on $2^{[n]}$ with mean measure Λ so that $\{X_a : a \in 2^{[n]}\}$ is a collection of independent Poisson random variables with each X_a having mean $\Lambda(a) \geq 0$. By ignoring multiplicities, each realization of this process defines a random subset $X^* := \{a \subseteq [n] : X_a > 0\} \in \mathcal{E}_n$ which consists of those points $a \in 2^{[n]}$ for which $X_a > 0$. The distribution of X^* is given by

$$\mathbb{P}(X^* \subseteq E) = \prod_{a \subseteq [n] : a \notin E} \exp\{-\Lambda(a)\}.$$

In general, X^* will not be monotone, but, as discussed above, it will have a least monotone cover given by $\alpha(X^*) := \{2^a : a \in X^*\} \in \mathcal{F}_n$.

It is straightforward to compute the induced distribution of $\alpha(X^*)$ on \mathcal{F}_n under the partial ordering

$\leq_{\mathcal{F}}$. That is, for $A \in \mathcal{F}_n$, let \bar{A} denote the complement of A in $2^{[n]}$, then

$$\begin{aligned}\mathbb{P}_n(\alpha(X^*) \leq_{\mathcal{F}} A) &= \prod_{a \in \bar{A}} \mathbb{P}(X_a = 0) \\ &= \prod_{a \in \bar{A}} \exp\{-\Lambda(a)\} \\ &= \exp\left\{-\sum_{a \in \bar{A}} \Lambda(a)\right\}.\end{aligned}$$

For each $n \geq 1$, let $\mathbf{0}_n^{\mathcal{E}}$ and $\mathbf{0}_n^{\mathcal{F}}$ be the minimal element of \mathcal{E}_n and \mathcal{F}_n respectively, i.e. $\mathbf{0}_n^{\mathcal{E}} = \mathbf{0}_n^{\mathcal{F}} = \emptyset$, and define $\mu_n([\mathbf{0}_n^{\mathcal{E}}, A]) := \mathbb{P}_n(X^* \leq_{\mathcal{E}} A) = \exp\{-\sum_{a \in [n]: a \notin A} \Lambda_n(a)\}$ to be the probability measure on \mathcal{E}_n and $\nu_n([\mathbf{0}_n^{\mathcal{F}}, B]) := \mathbb{P}(\alpha(X^*) \leq_{\mathcal{F}} B) = \exp\{-\sum_{b \in \bar{B}} \Lambda_n(b)\}$ the probability measure on \mathcal{F}_n induced by μ_n through α , shown above, for some non-negative mean measure $\Lambda_n(\cdot)$ on $2^{[n]}$.

For a subset $A \subset \mathbb{N}$, let $\#A$ denote the cardinality of A , i.e. the number of elements in A . We now show that a necessary and sufficient condition for μ_n , and hence ν_n , to be infinitely exchangeable under action of $(\varphi_{m,n}^{\mathcal{E}}, m \leq n)$ is that for every $n \geq 1$, $\Lambda_n(a) = \lambda_n(\#a)$ for some collection of non-negative real numbers $(\lambda_n(r), n \geq 1, 0 \leq r \leq n)$ which satisfy

$$\lambda_n(r) = \lambda_{n+1}(r) + \lambda_{n+1}(r+1). \quad (1)$$

Theorem 2.1. *The collection $(\mu_n, n \geq 1)$ of probability measures on $\mathcal{E}^{\text{rest}}$ and $(\nu_n, n \geq 1)$ on $\mathcal{F}^{\text{rest}}$ are infinitely exchangeable if and only if the collection of mean measures $(\Lambda_n, n \geq 1)$ satisfy*

- (a) *for every $n \geq 1$, $\Lambda_n(a) = \lambda_n(\#a)$ for all $a \in \mathcal{F}_n$ for some collection of measures $(\lambda_n(\cdot), n \geq 1)$ on the non-negative real numbers and*
- (b) *for every $n \geq 1$, $\lambda_n(r) = \lambda_{n+1}(r) + \lambda_{n+1}(r+1)$ for $r = 0, 1, \dots, n$.*

Moreover, if (a) and (b) hold, the measure $(\mu_n^*, n \geq 1)$ induced on $\mathcal{G}^{\text{rest}}$ by $(\mu_n, n \geq 1)$ through $\beta \circ \alpha$, i.e. $\mu_n^* := \mu_n \alpha^{-1} \beta^{-1} = \nu_n \beta^{-1}$, is infinitely exchangeable.

Proof. Suppose (a) and (b) hold. Let $n \geq 1$ and suppose $E \in \mathcal{E}_n$. Then

$$\mu_n([\mathbf{0}_n^{\mathcal{E}}, E]) = \exp\left\{\sum_{e \subseteq [n]: e \notin E} \lambda_n(\#e)\right\}.$$

Clearly, the μ_n are finitely exchangeable for each $n \geq 1$ as the measure depends only on the cardinality of the elements of $\{e \subset [n] : e \notin E\}$, which directly depends on the cardinality of the elements of E , which are invariant under permutations.

As noted in section 1.1, the pullback $\varphi_{m,n}^{\mathcal{E}-1}$ under the restriction maps takes intervals to intervals. Hence, for $E = \{E_1, \dots, E_k\} \in \mathcal{E}_m$, the interval $[\mathbf{0}_m^{\mathcal{E}}, E] \subseteq \mathcal{E}_m$ maps to the interval $[\mathbf{0}_n^{\mathcal{E}}, \sup \varphi_{m,n}^{\mathcal{E}-1}(E)] \subseteq \mathcal{E}_n$, where $\sup \varphi_{m,n}^{\mathcal{E}-1}(E) = E \cup \{m+1, \dots, n\}$ is the unique maximal element of $\varphi_{m,n}^{\mathcal{E}-1}(E)$ in \mathcal{E}_n . To simplify notation, write $\sup_{m,n} E := \sup \varphi_{m,n}^{\mathcal{E}-1}(E)$ in what follows.

Hence, for consistency under sampling by restriction maps we must have

$$\mu_n([\mathbf{0}_n, E]) = \mu_{n+1}([\mathbf{0}_{n+1}, \sup_{n,n+1} E]) \quad (2)$$

Consistency follows from this since we now have

$$\mu_{n+1}([\mathbf{0}_{n+1}, \sup_{m,n} E]) = \exp \left\{ \sum_{e \subseteq [n+1]: e \notin \sup_{m,n} E} \lambda_{n+1}(\#e) \right\},$$

which reduces (2) to

$$\sum_{e \subseteq [n]} \lambda_n(\#e) = \sum_{e' \subseteq [n+1]: e' \notin \sup_{m,n} E} \lambda_{n+1}(\#e') \quad (3)$$

which is a sum over subsets of the power set of $[n]$ and $[n+1]$ respectively. Under restriction, each $e \in 2^{[n]}$ corresponds to a two element subset of $[n+1]$, namely $\{e, e \cup \{n+1\}\}$. Hence, for $\{e \subseteq [n] : e \notin E\} := \{e_1, \dots, e_k\}$, we have $\{e' \subseteq [n+1] : e' \notin \sup_{m,n} E\} = \{e_j, e_j \cup \{n+1\} : j = 1, \dots, k\}$ and (3) is just

$$\sum_{e \subseteq [n]: e \notin E} \lambda_n(\#e) = \sum_{e \subseteq [n]: e \notin E} [\lambda_{n+1}(\#e) + \lambda_{n+1}(\#e + 1)].$$

Infinite exchangeability for ν_n is endowed by μ_n through α , as discussed in section 1.6.

For the reverse implication, note that we start with the condition on the mean measures Λ_n

$$\sum_{e \subseteq [n]: e \notin E} \Lambda_n(e) = \sum_{e \subseteq [n]: e \notin E} [\Lambda_{n+1}(e) + \Lambda_{n+1}(e \cup \{n+1\})]$$

and exchangeability requires $\Lambda_n(a) = \Lambda_n(b)$ for all $a, b \subset [n]$ with $\#a = \#b$ so we can reduce this to the collection of measures $\lambda_n(\cdot)$ taking values in the non-negative real numbers, as we have done above. And consistency requires of either μ_n or ν_n requires (b) to hold.

The infinite exchangeability of the measures $(\mu_n^*, n \geq 1)$ on $\mathcal{G}^{\text{rest}}$ is a direct corollary of the infinite exchangeability of μ_n and the action of the functor $\beta \circ \alpha : \mathcal{E}^{\text{rest}} \rightarrow \mathcal{G}^{\text{rest}}$. \square

Corollary 2.2. *Suppose $(\lambda_n(r), n \geq 1, r = 0, \dots, n)$ is a doubly indexed sequence of non-negative real numbers satisfying (1), then there exists a measure μ^* on \mathcal{G} , the space of graphs with vertex set \mathbb{N} such that*

$$\mu_n^*(G) = \mu^* \left(\left\{ G^* \in \mathcal{G} : G_{[[n]]}^* = G \right\} \right),$$

where $G_{[[n]]}^*$ denotes the restriction of G^* to the vertex set $[n]$ and μ_n^* are the induced measures on \mathcal{G}_n given above.

We make note of a correspondence between the solutions to (1) and the classical Hausdorff moment problem. Connection between the Hausdorff moment problem and de Finetti's theorem have been shown by Diaconis and Freedman [8] and are well known throughout the literature. Some usable choices for $(\lambda_n(r), 0 \leq r \leq n)$ are

- $\lambda_n(r) \propto \alpha^r (1 - \alpha)^{n-r}$ for $0 < \alpha < 1$ and
- $\lambda_n(r) \propto \binom{n}{r}^{-1}$.

3 Cluster analysis

Given a measure μ_n^* on \mathcal{G}_n which is based on a collection $(\lambda_n(r), n \geq 1, r = 0, 1, \dots, n)$ which satisfy (1), we can easily calculate the marginal distribution that a triple of vertices, e.g. i, j, k , is transitive, i.e. $i \sim j, i \sim k, j \sim k$, given two are adjacent, e.g. $i \sim j$ and $i \sim k$, by a standard exchangeability argument. In particular, for any $n \geq 3$ and $i, j, k \in [n]$ we have

$$\begin{aligned} \mu_n^*[\{i \sim j, i \sim k, j \sim k\} | \{i \sim j, i \sim k\}] &= \mu_3^*[\{1 \sim 2, 1 \sim 3, 2 \sim 3\} | \{1 \sim 2, 1 \sim 3\}] \\ &= \frac{1 - e^{-\lambda_3(3)}(1 - (1 - e^{-\lambda_3(2)})^3)}{1 - e^{-\lambda_3(3)}(1 - (1 - e^{-\lambda_3(2)})^2)}, \end{aligned}$$

which, unlike the Erdős-Rényi process, does not correspond to the clustering coefficient. This calculation only accounts for the subgraph comprised of the vertices i, j, k and any edges between them. In general, if we observe a graph $G \in \mathcal{G}_n$ and we are told that $i \sim j$ and $i \sim k$ but not told if there is an edge between j and k or not, but we also know the rest of the graph structure, this information is relevant for determining whether or not $j \sim k$ in G , and has implications in the inference of missing links in network data sets.

For example, consider the two graphs G_1 and G_2 below where

$$G_1 = \begin{matrix} & \begin{matrix} 1 & 2 & 3 & 4 \end{matrix} \\ \begin{matrix} 1 \\ 2 \\ 3 \\ 4 \end{matrix} & \begin{pmatrix} - & & 1 & 0 \\ & - & 1 & 0 \\ 1 & 1 & - & 1 \\ 0 & 0 & 1 & - \end{pmatrix} \end{matrix}$$

and

$$G_2 = \begin{matrix} & \begin{matrix} 1 & 2 & 3 & 4 \end{matrix} \\ \begin{matrix} 1 \\ 2 \\ 3 \\ 4 \end{matrix} & \begin{pmatrix} - & & 1 & 1 \\ & - & 1 & 1 \\ 1 & 1 & - & 0 \\ 1 & 1 & 0 & - \end{pmatrix} \end{matrix},$$

and the presence or absence of an edge between 1 and 2 is unknown, but the rest of this network is known. Given what we do know about G_1 , there are three possible monotone sets which correspond to G_1 ,

$$\langle \{1, 2\}, \{1, 3\}, \{2, 3\}, \{3, 4\} \rangle, \quad \langle \{1, 2, 3\}, \{3, 4\} \rangle, \quad \langle \{1, 3\}, \{2, 3\}, \{3, 4\} \rangle.$$

On the other hand, there are five monotone sets corresponding to G_2 ,

$$\begin{aligned} &\langle \{1, 2, 4\}, \{1, 2, 3\} \rangle, & \langle \{1, 2, 3\}, \{1, 3\}, \{2, 3\} \rangle & \quad \langle \{1, 2, 3\}, \{1, 4\}, \{2, 4\} \rangle \\ &\langle \{1, 2\}, \{1, 4\}, \{2, 4\}, \{2, 3\}, \{1, 3\} \rangle, & \langle \{1, 3\}, \{1, 4\}, \{2, 3\}, \{2, 4\} \rangle. \end{aligned}$$

So the information given by the rest of the network, and any edges, or lack thereof, which involve any of the i, j, k of interest, affects the conditional probability of $j \sim k$, e.g. $1 \sim 2$ in this case.

In general, the clustering of this process is expected to be larger than the marginal probability expression above as the presence of a cluster of three vertices increases the probability of other clusters which involve these vertices. The nature of the construction, e.g. description of the functor $\beta \circ \alpha$, leads to overlapping of the various subsets of A which is “forgotten” in the projection onto \mathcal{G}_n and provides various ways by which clustering can occur.

The construction also allows us to move between the space of graphs and that of monotone sets, which is helpful in calculations.

3.1 Detecting clusters

The nature of this construction naturally lends itself to methods in cluster analysis, which has been studied in certain applications in statistics and machine learning [2, 10, 11]. The setting is as follows. Let $n \geq 1$ be the size of a sample for which we label statistical units, e.g. individuals, arbitrarily in $[n]$ and observe a network for this sample, i.e. an undirected graph $G \in \mathcal{G}_n$. Along with $[n]$, let \sim_1, \dots, \sim_k be a collection of different equivalence relations on $[n]$. A collection of labels $\{i_1, \dots, i_m\}$ is said to form a *cluster*, or community, in our network if, for some $l = 1, \dots, k$, $i_p \sim_l i_q$ for every $p, q = 1, \dots, m$. Inferring clusters in networks has implications, for example, in the problem of data deduplication and parsing for semi-structured text data sets as well as inferring communities and missing links in social networks.

In a statistical setting for social networks could represent different ‘types’ of relationships among individuals. That is, individuals in a social network are associated by certain relationships which underlie the network, e.g. a 4-node clique in a network could arise from 4 nodes belonging to the same cluster, or due to the overlap of two 3-node clusters and a 2-node cluster, e.g. $\{1, 2, 3\}, \{1, 2, 4\}$ and $\{3, 4\}$, which both result in the presence of the clique $\{1, 2, 3, 4\}$ in the projected network.

3.1.1 A statistical model

In the setting of section 2 consider an infinite population \mathcal{U} of units from which we sample a finite number $n \geq 1$ which we label in $[n]$, i.e. our sample is u_1, \dots, u_n , and we observe for this sample a network, or graph, $G \in \mathcal{G}_n$ which we assume to have been generated according to the Poisson point process (PPP) construction on $2^{[n]}$ which we laid out above. In particular, let $\Lambda := (\Lambda_n(\cdot), n \geq 1)$ be a family of mean measures on $(2^{[n]}, n \geq 1)$ such that $\Lambda_n(a) = \lambda_n(\#a)$ for every a and the $\lambda_n(\cdot)$ satisfy (1). Let μ_n, ν_n and μ_n^* be the measures on $\mathcal{E}_n, \mathcal{F}_n$ and \mathcal{G}_n defined in above sections. Then we have the following model:

$$\begin{aligned} X &\sim \text{PPP}(\Lambda) \\ G|\Lambda &\sim \mu^*. \end{aligned}$$

We now imagine reversing this process to infer whether a given complete subgraph $H \subset G$ of units $\{u_1, \dots, u_h\}$ represents a cluster of u_1, \dots, u_n . This amounts to computing the conditional probability that $X_H > 0$ given that $H \subset G$.

Suppose we observe a network $G \in \mathcal{G}_n$ with complete subgraph H . Under the inverse image of the functor β , we have that $\beta^{-1}(G) \in \mathcal{F}_n$ is a collection of monotone sets which correspond to G . Furthermore, the inverse image of the least monotone cover $\alpha^{-1}[\beta^{-1}(G)]$ is a collection of possible subsets of $2^{[n]}$ which have least monotone covers corresponding to $\beta^{-1}(G)$. As we have shown in theorem 2.1, the consistency condition in (1) guarantees infinite exchangeability of (μ_n) as well as $(\mu_n \alpha^{-1})$ and $(\mu_n(\beta\alpha)^{-1})$ on $\mathcal{F}^{\text{rest}}$ and $\mathcal{G}^{\text{rest}}$ respectively.

Hence $(\beta\alpha)^{-1}(G)$ is the collection of random subsets in \mathcal{E}_n which correspond to G . Define $Z := Z(H, G) \subseteq (\beta\alpha)^{-1}(G)$ to be the elements of $(\beta\alpha)^{-1}(G)$ which contain the set H , and \bar{Z} the complement of Z in $(\beta\alpha)^{-1}(G)$, i.e. $\bar{Z} := (\beta\alpha)^{-1}(G) - Z$. Then the conditional probability that H is a cluster in the subsample of \mathcal{U} given G is

$$\pi(H; G) := \mathbb{P}(H \text{ a cluster}) = \mu_n(Z) / \mu_n[(\beta\alpha)^{-1}(G)]. \quad (4)$$

The conditional probability calculation in (4) is valid for inference of a cluster H if we are interested in the presence specifically of a cluster with exactly the elements of H and not a cluster $H^* \supset H$ which contains the entire cluster H . In applications, we might not be able to observe clusters at such a fine level. In particular, if there is a cluster $H^* \supseteq H$ in the population, we may not be able to observe the presence or absence of a sub-cluster H within the larger cluster. In this case, it is reasonable to consider performing inference at a coarser level. That is, for a given collection $H := \{u_1, \dots, u_h\}$, we wish to determine whether there is a cluster $H^* \supseteq H$ in the population.

Given a graph with clique H , let $H^*(H, G) \subseteq \beta^{-1}(G)$ be the monotone subset of $\beta^{-1}(G)$ such that $H \in H^*$. In particular,

$$H^*(H, G) := [H, \mathbf{1}_n] \cap \beta^{-1}(G).$$

Given G , the conditional probability of $H^*(H, G)$ is

$$\mathbb{P}(H^*(H, G)|G) = \nu_n [[H, \mathbf{1}_n] \cap \beta^{-1}(G)] / \nu_n [\beta^{-1}(G)].$$

3.1.2 Stochastic classification

Stochastic classification models have been studied previously by McCullagh and Yang [11] in the context of the Gauss-Ewens clustering process. In that model, a finite sample u_1, \dots, u_n is taken from an infinite population \mathcal{U} of units for which we observe some feature $Y_i := Y(u_i) \in \mathcal{S}$, for some subspace \mathcal{S} , usually $\mathcal{S} \subseteq \mathbb{R}^d$. Associated to \mathcal{U} is a partition B of \mathbb{N} and conditional on B the vector $Y := (Y_1, \dots, Y_n)$ is normally distributed with mean and covariance which depend on B only through the restriction $B|_{[n]}$ of B to $[n]$. For a newly sampled individual u^* , it is shown how to classify u^* based on its feature $Y^* := Y(u^*)$ and the data (Y_1, \dots, Y_n) and $B|_{[n]}$ already obtained. This is carried out by computing the conditional distribution that u^* belongs to each block $b \in B|_{[n]}$ or a possibly new block of B given the data Y_1, \dots, Y_n, Y^* and $B|_{[n]}$. The infinite exchangeability of the Gauss-Ewens process is a tool which allows the computation of conditional distributions in this setting.

In our setting, we also assume an infinite population \mathcal{U} and an associated network (undirected graph) G of \mathcal{U} which is generated by the Poisson point process recipe above. Suppose we sample u_1, \dots, u_n from \mathcal{U} and we observe the restriction $G|_{[n]}$ of G to $[n]$ as well as the component of \mathcal{E}_n which corresponds to $G|_{[n]}$. That is, we assume we observe the realization of X^* , the projection of the Poisson point process on $2^{[n]}$ onto \mathcal{E}_n . For a new individual $u^* \in \mathcal{U}$, suppose we observe its connections within the network, i.e. we have complete knowledge of $G|_{[n+1]}$, but nothing about its associated clusters in \mathcal{E}_{n+1} .

Infinite exchangeability of the measures $(\mu_n, n \geq 1)$ on $\mathcal{E}^{\text{rest}}$ makes the specification of the conditional distribution of various outcomes straightforward. Given $X^* \in \mathcal{E}_n$ and $\tilde{G} \in \mathcal{G}_{n+1}$, let $H^* \in \mathcal{E}_{n+1}$ be a collection which is consistent with X^* and \tilde{G} , i.e. $H^* \subseteq (\beta\alpha)^{-1}(\tilde{G}) \cap \varphi_{n,n+1}^{\mathcal{E}^{-1}}(X^*)$, we have

$$\text{pr}(H^*|\tilde{G}, X^*) = \frac{\mu_{n+1}(H^*)}{\mu_{n+1}[(\beta\alpha)^{-1}(\tilde{G}) \cap \varphi_{n,n+1}^{\mathcal{E}^{-1}}(X^*)]}.$$

References

- [1] Awodey, S. (2006). *Category Theory*. Oxford: Clarendon Press.
- [2] Blei, D., Ng, A. and Jordan, M. (2003). Latent Dirichlet allocation. *J. Machine Learning Research* **3**, 993–1022.

- [3] Bollobás, B. and Riordan, O. (2004). Diameter of a scale free random graph. *Combinatorica* **24**, 5–34.
- [4] Bollobás, B. and Riordan, O. (2011). Sparse Graphs – Metrics and Random Models. *Random Structures and Algorithms* **39**, 1–38.
- [5] Bollobás, B., Janson, S. and Riordan, O. (2007). The phase transition in inhomogeneous random graphs. *Random Structures and Algorithms* **31**, 3–122.
- [6] Bollobás, B., Janson, S. and Riordan, O. (2011). Sparse random graphs with clustering. *Random Structures and Algorithms* **38**, 269–323.
- [7] Broderick, T., Jordan, M. and Pitman, J. (2011). Beta processes, stick breaking, and power laws. *Bayesian Analysis*, to appear. Available at arXiv:1106.0539.
- [8] Diaconis, P. and Freedman, D. A. (2004). The Markov moment problem and de Finetti’s theorem: Parts I and II. *Mathematische Zeitschrift*, **247**, 183–212.
- [9] McCullagh, P. (2002). What is a statistical model? *Annals of Statistics* **30**, 1225–1310.
- [10] McCullagh, P. and Yang, J. (2006). Stochastic classification models. *Proc. Int. Congress of Mathematicians, Madrid III*, 669–686.
- [11] McCullagh, P. and Yang, J. (2008). How many clusters? *Bayesian Analysis* **3**, 101–120.
- [12] Pitman, J. (2005). *Combinatorial Stochastic Processes*. Springer, New York.